EDA Case Study

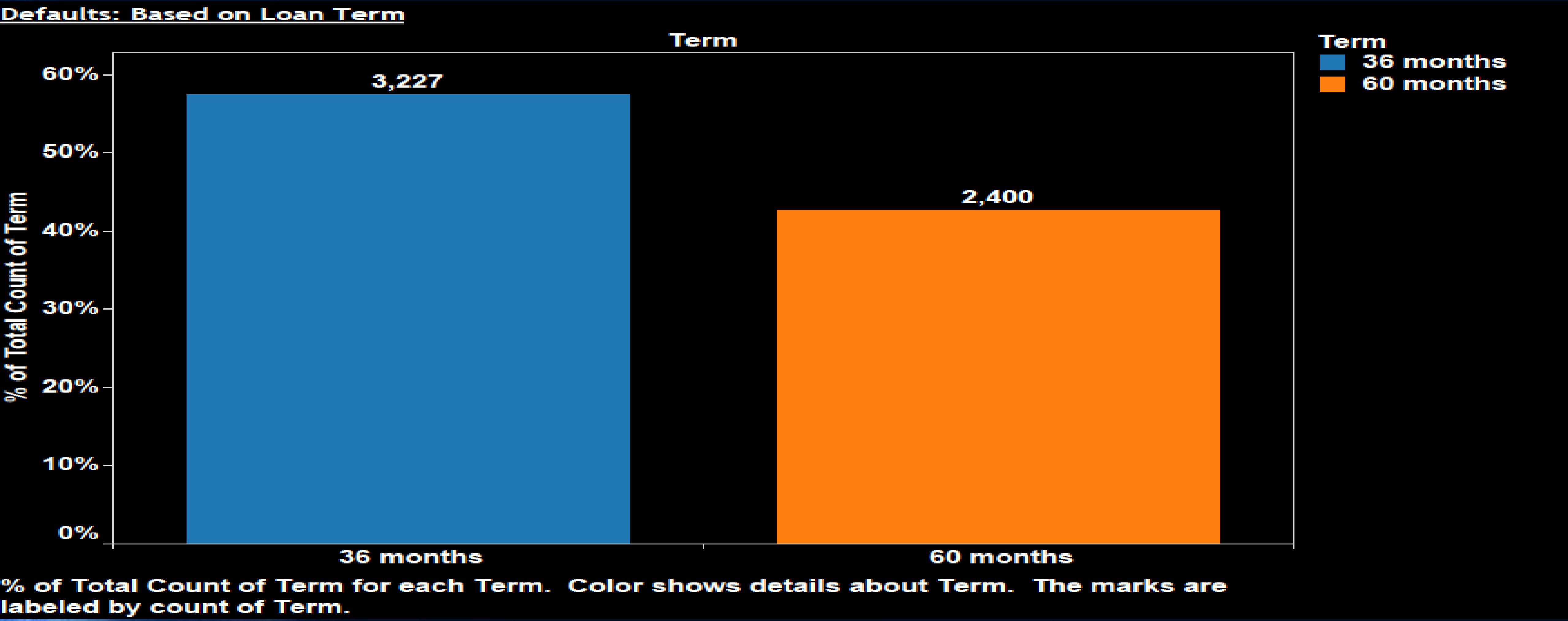
EXPLORATORY DATA ANALYSIS:

- AISHWARYA
- NAVEEN
- AJAY
- CHETAN

Problem Statement & Objective

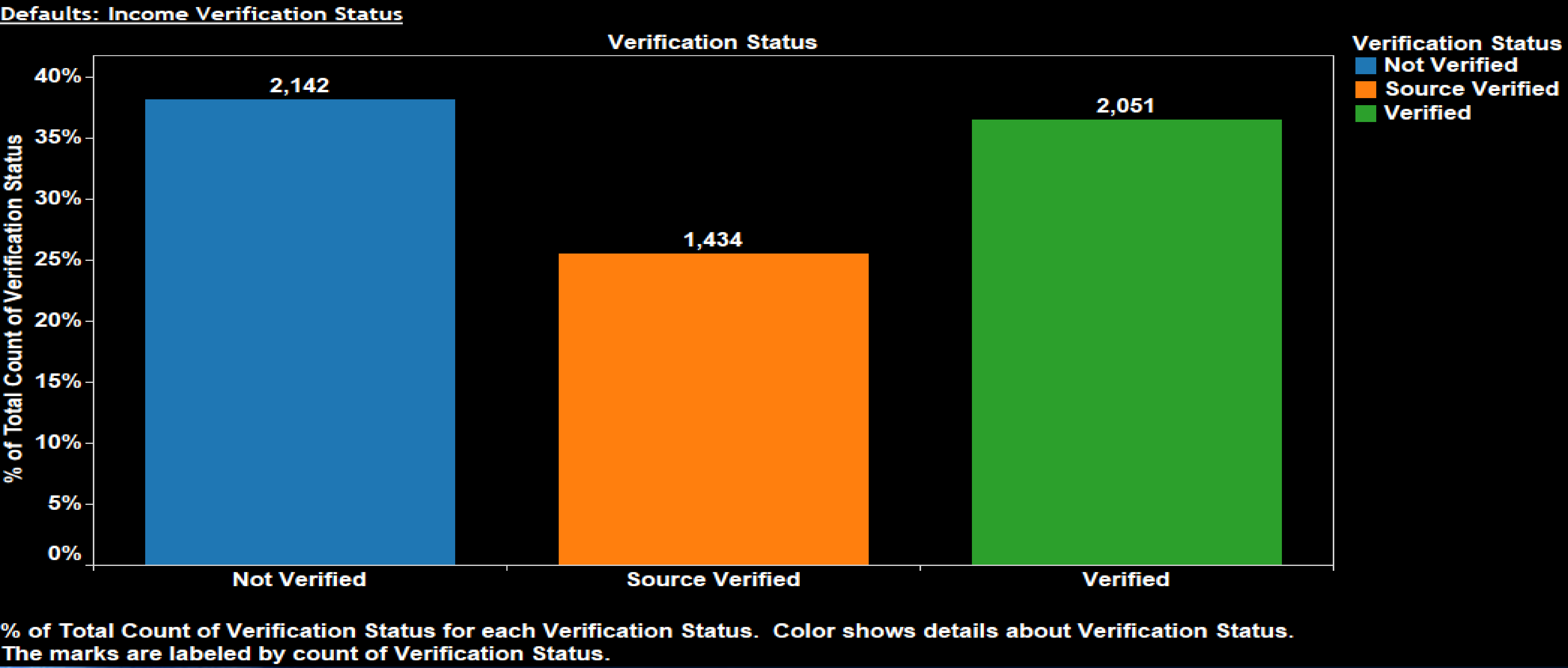
- Lending loans to “Risky” applicants is the major reason for credit loss (“Charged Off” cases)
- If we can identify such “Risky” applicants, bad loans can be reduced significantly and thereby credit loss can bedrastically reduced.
- Endeavor is to analyze historical information about “Charged Off” cases and identify driving factors behind such loans defaults.
- Company wish to use this insight for an effective portfolio and risk assessment.
- Exploratory Data Analysis is to be performed to get more insight.

Loan Term



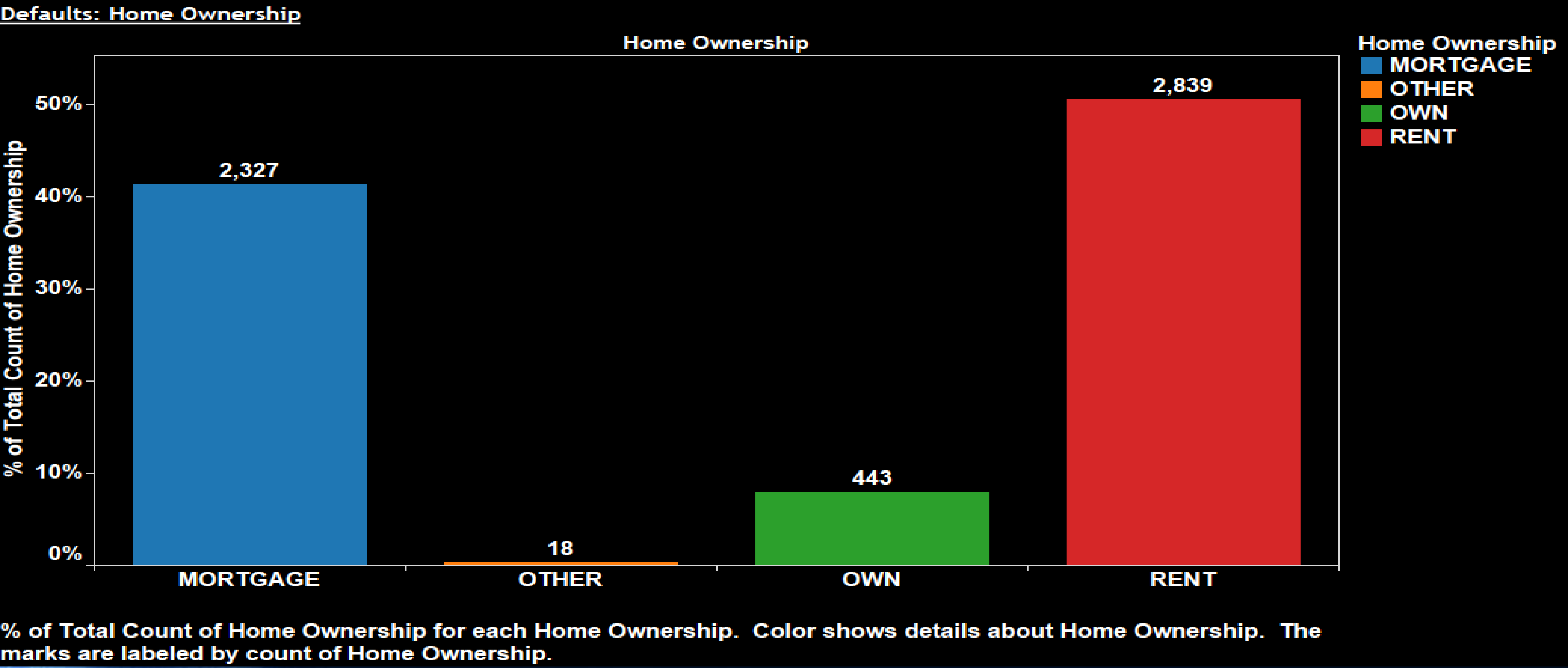
- Term of the loan does not seem to impact on defaults.

Customer / Income Source Verification Status



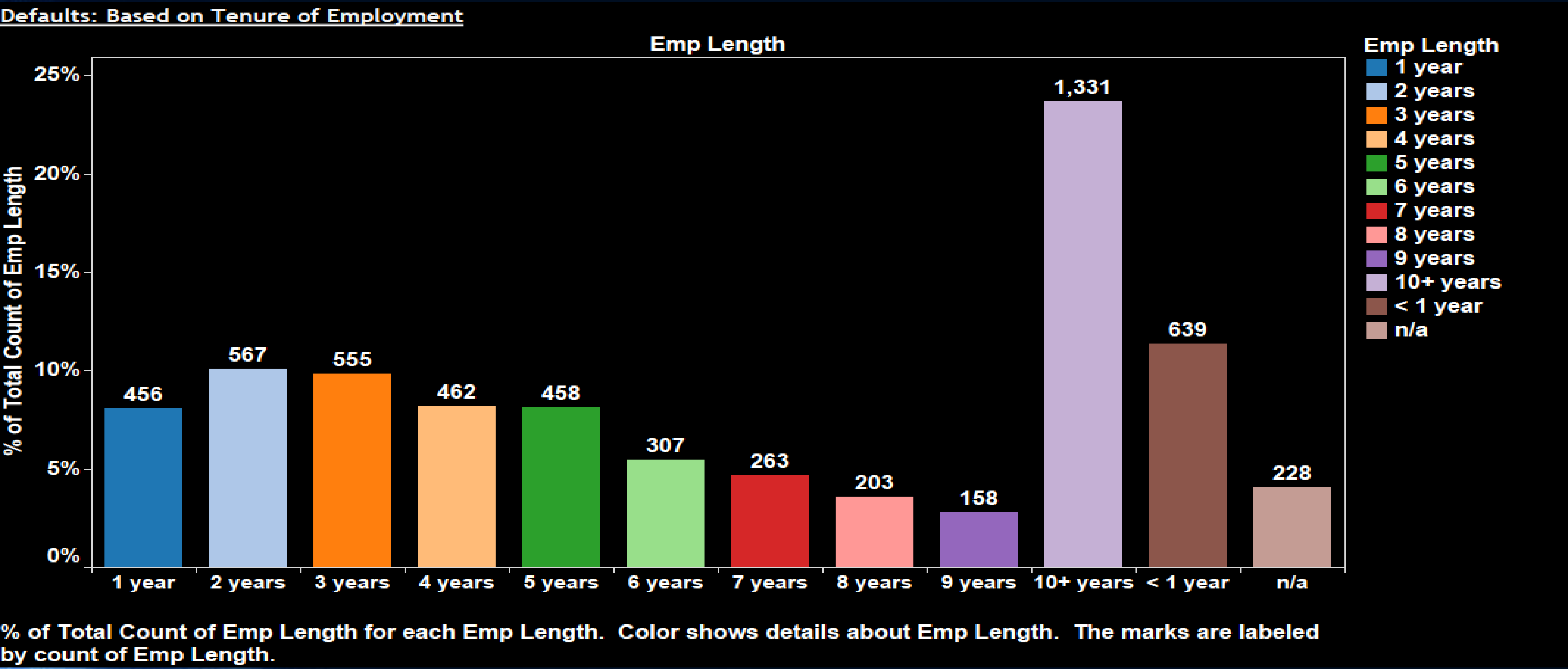
- Default of loans is not significantly effected by the fact that customer and it's source of income are verified or not.

Status of Home Ownership



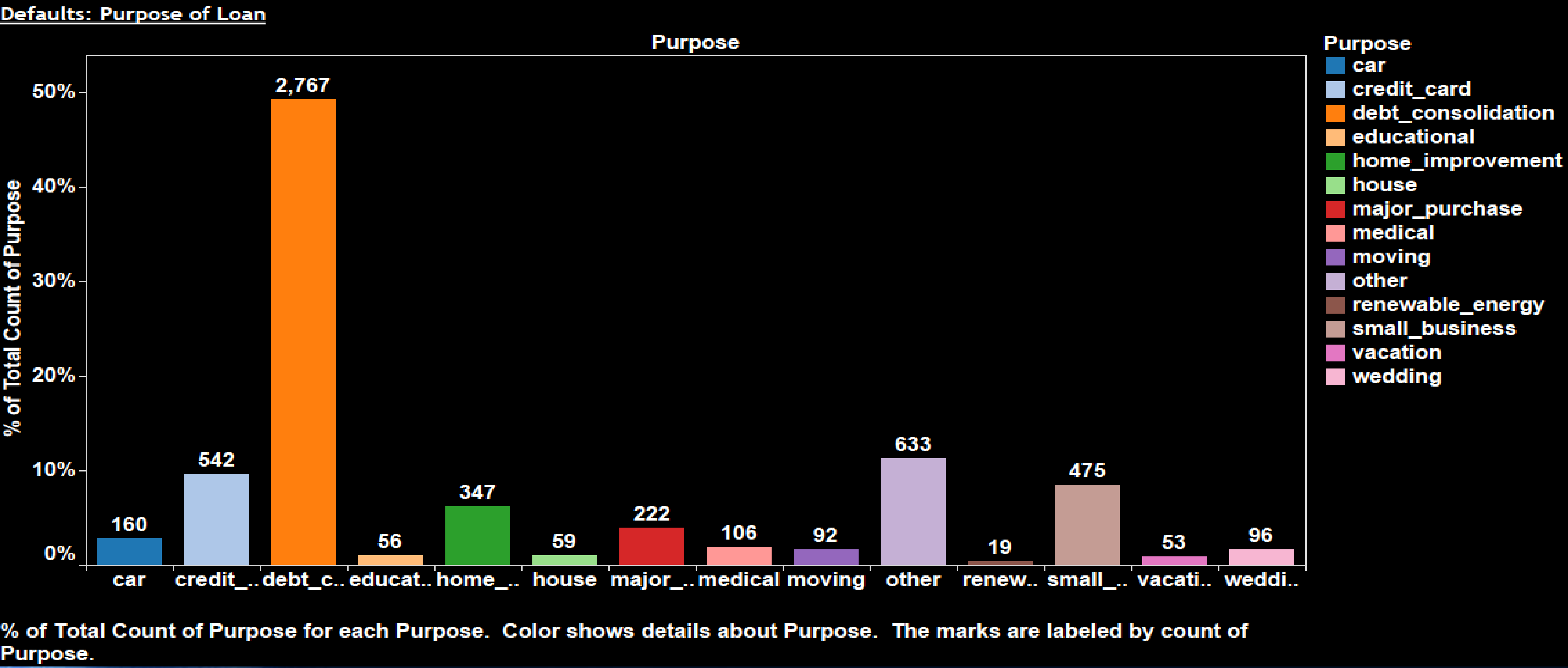
- It is established that customers having rented or mortgaged homes are more prone to defaults as compared to customers having no home related obligations.

Tenure of Employment Length



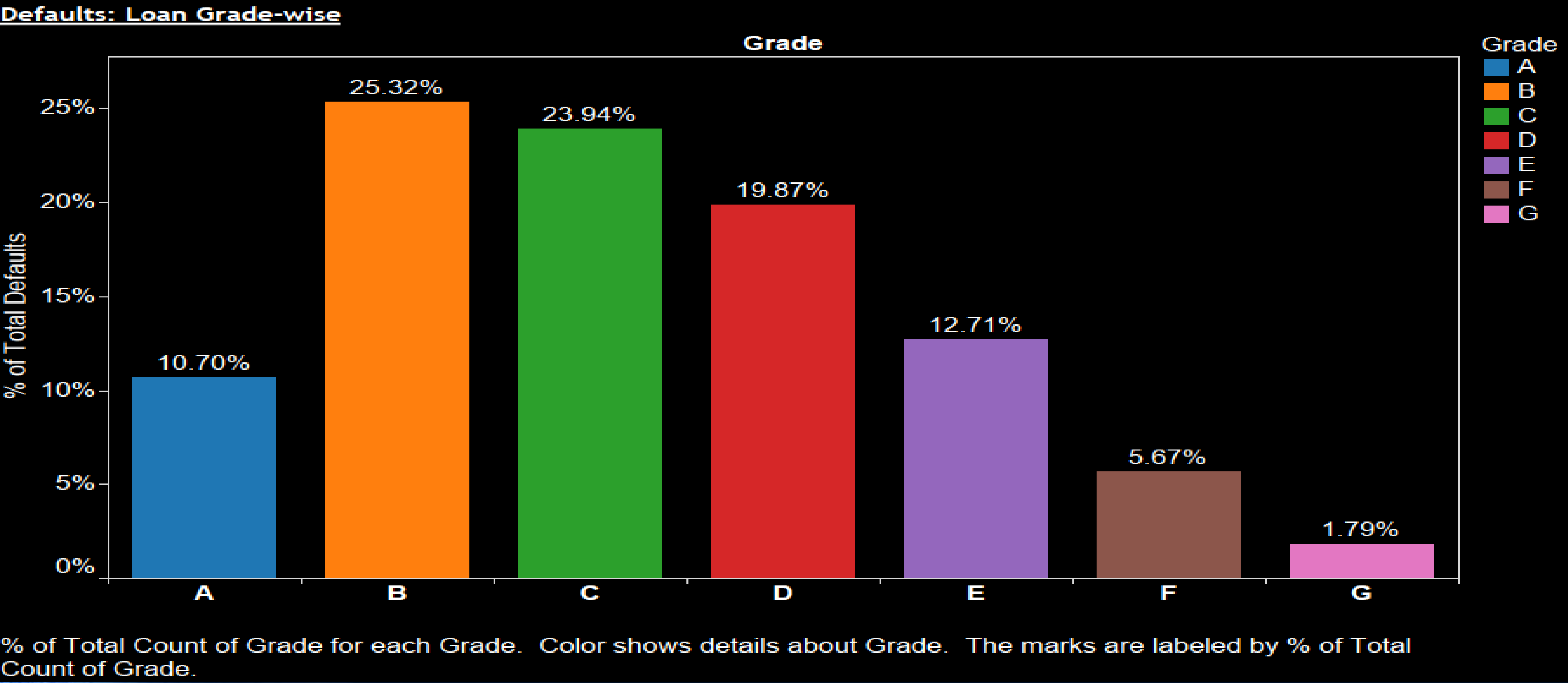
- Interestingly defaults are quite significant where tenure of employment is either 10+ years or less than a year.

Purpose of Loan



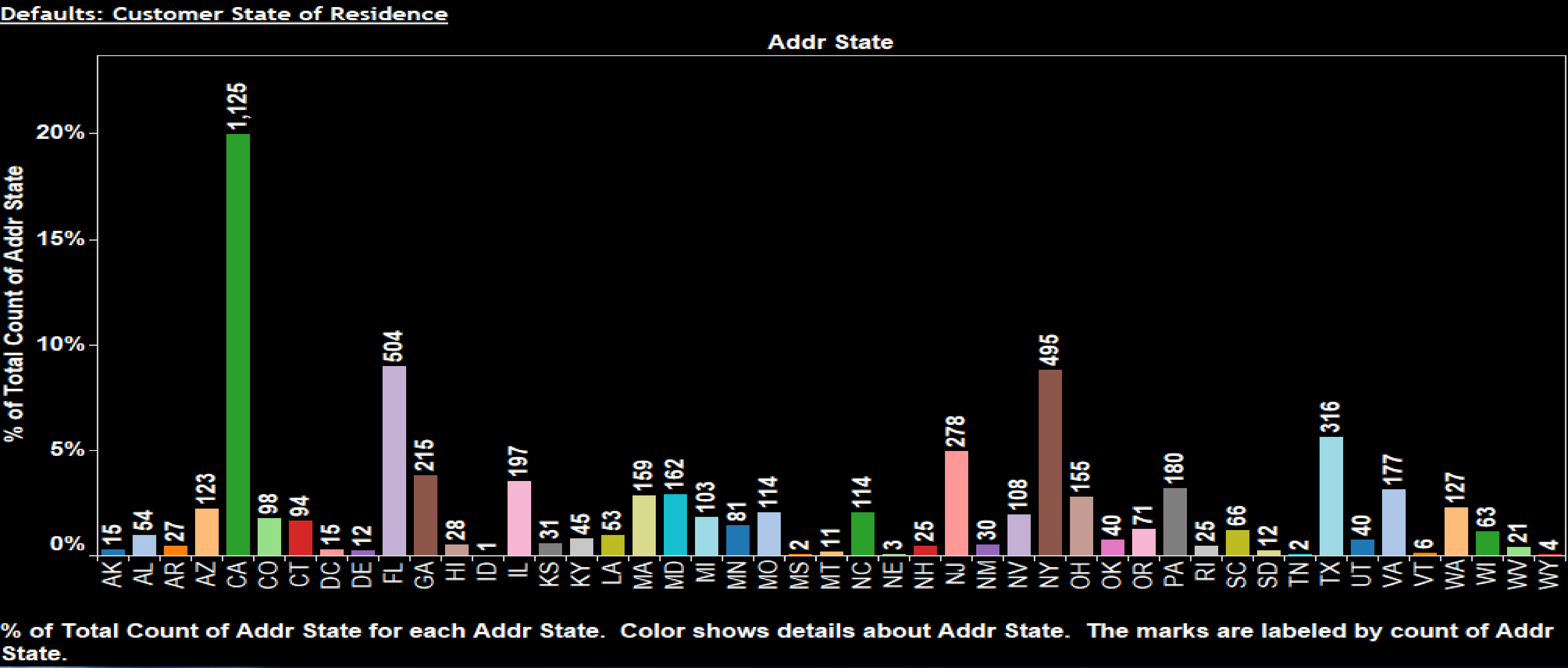
- Debt Consolidation is the major purpose of loan contributing to defaults.

Loan Grade



- Loan Grade B, C & D shows more defaults as compared to others.

State of Residence of Default Customers



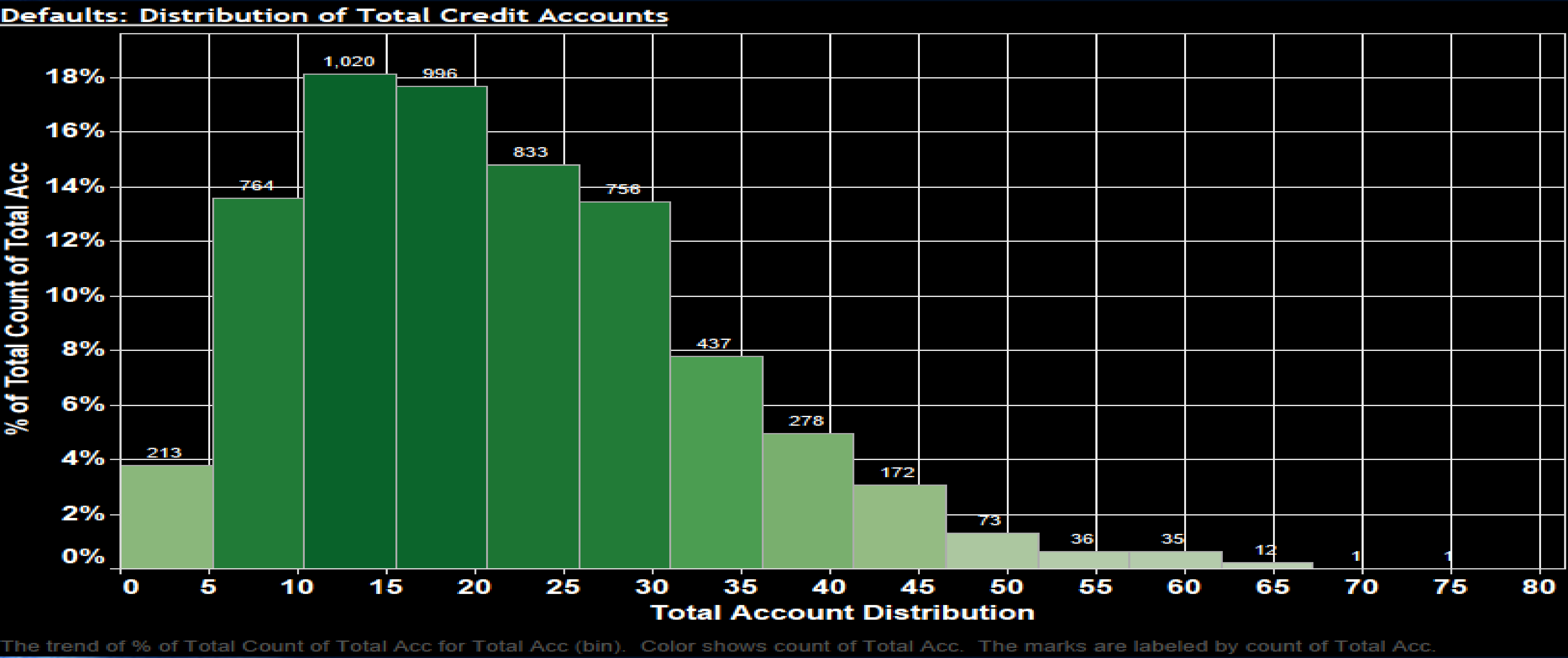
- Default customers are majorly from CA, FL, NY/NJ and TX States as compared to others.

Customer Income Class

% of Total Count of Class for each Class. Color shows details about Class. The marks are labeled by count of Class. The data is filtered on Loan Status, which keeps Charged Off.

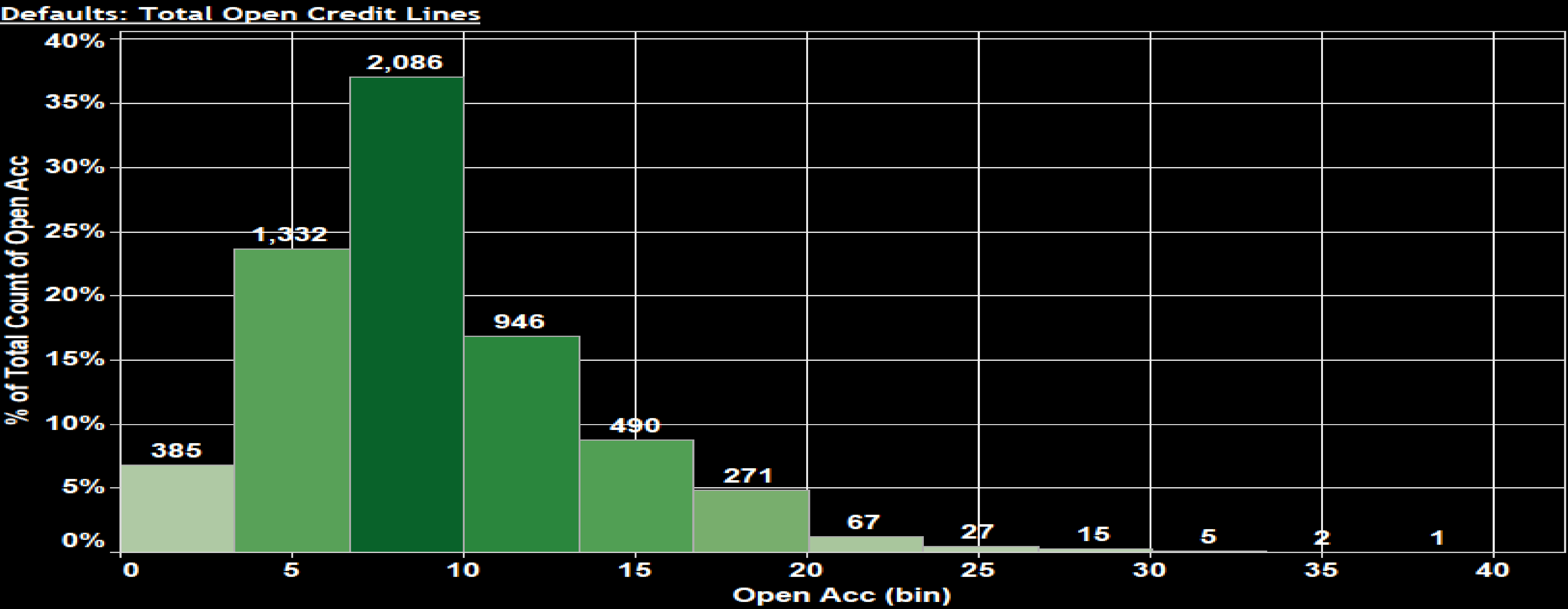
- Customers in a Middle Income Class tend to default more.

Distribution of Credit Accounts



- Generally defaults are linked to more number of credit lines.

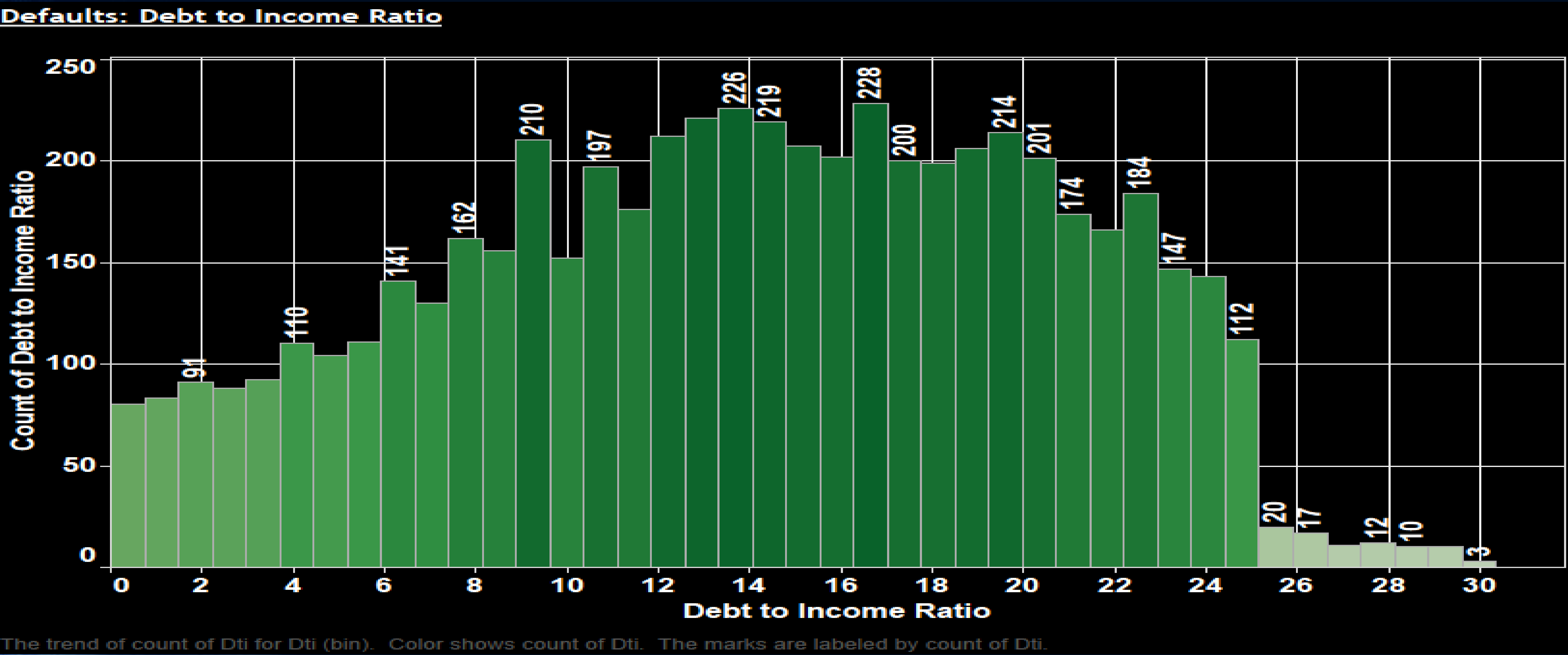
Total Open Credit Lines



The trend of % of Total Count of Open Acc for Open Acc (bin). Color shows sum of Open Acc. The marks are labeled by count of Open Acc.

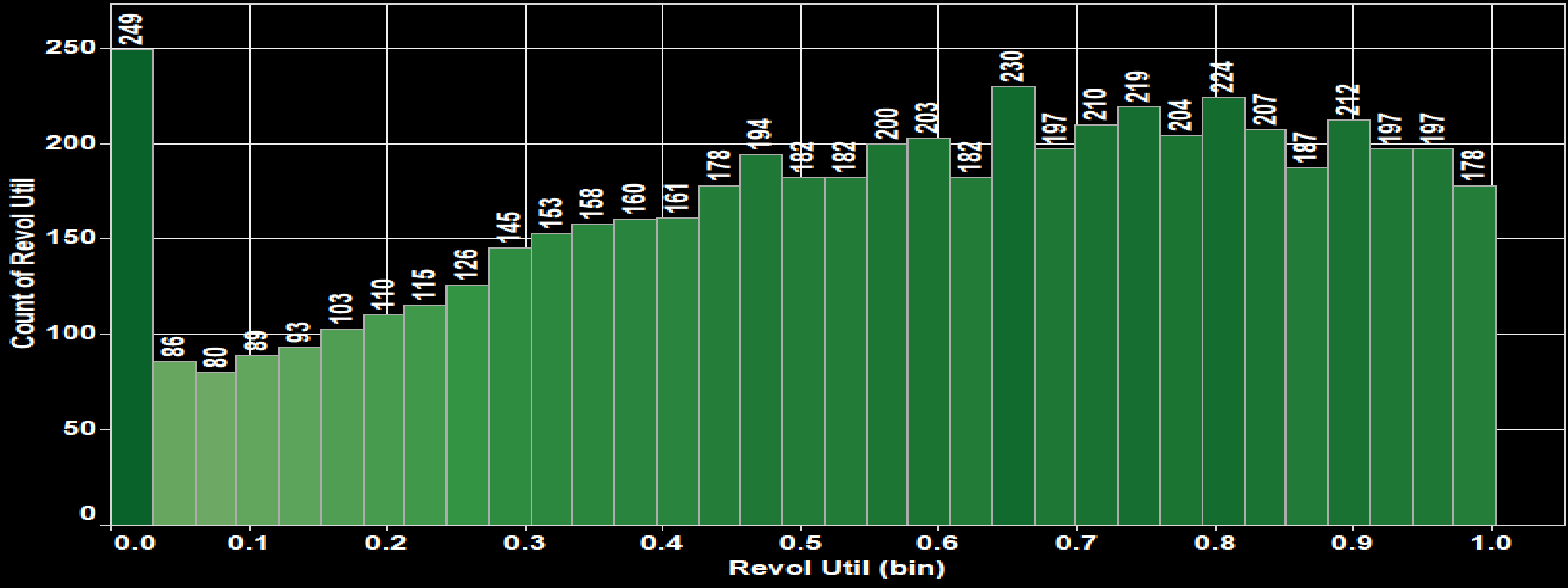
- Higher the Total Open Credit Lines with customer, higher the chances of defaults.

Customer Debt to Income Ratio (DTI)



- More the DTI, more chances of customer default.

Revolving Utilization



The trend of count of Revol Util for Revol Util (bin). Color shows count of Revol Util. The marks are labeled by count of Revol Util.

- In general, higher revolving utilization results into default.

Conclusion:

1. Term of Loan and Verification of Income Source do not have any relation with default of loans.
2. Customers having home obligations like mortgage or paying rent are more prone to defaults.
3. Customer who are very new in employment (< 1 year) or who are employed for a long time (> 10 years) result into more default of loans.
4. Loan taken for Debt Consolidation are more likely to get defaulted.
5. B, C and D types of loan are major loan type getting defaulted as compared to other types.
6. Customer residents in California, Florida, Texas, New York and New Jersey are higher in defaults as compared to customers in other states.
7. Customers in Middle Income Class results into more defaults as compared to other Lower and Upper Middle Class.
8. Higher the total credit accounts, total open accounts, DTI or Revolving Utilization, higher the chances of customer defaulting the loans.

Thank You….